Oxford Commission on AI & Good Governance

OXFORD INTERNET INSTITUTE

UNIVERSITY OF OXFORD

AI in the Public Service: From Principles to Practice

Oxford Commission on AI & Good Governance

DECEMBER 2021

AI in the Public Service: From Principles to Practice

Oxford Commission on AI & Good Governance

DECEMBER 2021

Oxford Commission on AI & Good Governance

Contents

Executive Summary 1

Foreword 2

1. Introduction 3

2. Challenges 4

3. Principles 7

4. Recommendations 8

    Who Should Provide Guidance on Public Service AI? 8

        Scientific Advisory Body on Public Impacts 8

        Independent and Accessible Arbitral Body 9

        Working Together 10

        Recommendation 1 10

    How Can We Build Public Service Capacities for AI for Good Governance? 11

        Recommendation 2 11

    How Can We Ensure Public Service AI is Trustworthy and Trusted? 12

        Recommendation 3 12

5. Conclusion: Ideas for Immediate Action 13

Objectives Summary 14

6. OxCAIGG Research 15

About the Commission 15

Commissioners 16

Acknowledgements 18

Executive Summary

Early experience indicates that there will be serious challenges as governments around the world apply artificial intelligence in public service.

This is the moment of opportunity, however, in which cooperation can ensure that AI systems are used for good governance and help us address some of the most pressing and intractable public problems.

Through research and deliberation, the Oxford Commission on AI and Good Governance reviewed the key challenges to ensuring that AI systems are effectively used for public service and arrived at three critical *questions*:

- Who should provide guidance on public service AI?
- How can we build public service capacities for AI for good governance?
- How can we ensure public service AI is trustworthy and trusted?

We affirmed that for AI to be used in the public interest, several *objectives* must be met:

- The design of AI must be inclusive.
- Any procurement of AI must be guided by an informed public agency.
- The implementation of AI in public service must be purposeful.
- AI systems must be persistently accountable to stakeholders.

To achieve these objectives, we make three *recommendations*:

I. Internationally, governments, industry, and civil society must work together to build and empower (a) an international **scientific body** for advancing research on public service AI applications and (b) an **arbitral body** for adjudicating disputes that might arise between the stakeholders involved with public service AI systems.

2. National governments, taking advantage of support from these two new international organisations, must (a) **build the capacity** of their public service to be deeply engaged with the design, procurement, implementation, and accountability of public service AI systems and (b) provide **toolkits for the staff** of public agencies to do their supervisory work.

3. Trust in the use of AI must be strengthened through public education campaigns about the everyday applications and real use cases on the horizon, their impacts, and their risks, by (a) having governments **openly disclose** how AI technologies are being used in the public service and (b) introducing a multisector agency which **provides a basic certification** system that continually validates applications and builds trust over time.

Progress towards the objectives can be assisted by three immediate actions:

I. Conducting feasibility studies for both the scientific and arbitral agencies.

II. Consulting with existing national and multilateral agencies on how these bodies could extend and supplement current capacities and best practices.

III. Planning for engagement at several of the important milestone events on technology innovation already on the international calendar over the next 36 months.

Foreword

Artificial intelligence promises to solve some of the most pressing challenges of public policy and societal change, but our governments might not be ready to deploy it for good governance.

In recent years, many governments have sought to take advantage of new technologies like machine learning, big data, and other algorithmic tools to develop and implement better policies and programmes for their constituents. While public sector innovation is to be encouraged, many of these projects have struggled to get off the ground, collaborations with industry have broken down, and poor-quality data and flawed technologies have resulted in unexpected and often biased outcomes.

These experiences have raised a very important question: how can AI be best harnessed for public service?

From 2020 to 2021 we participated in a project, chartered by the University of Oxford, to answer this very question—the Oxford Commission on AI and Good Governance (OxCAIGG). Our group included independent directors, executives, academics, lawyers, and government advisers representing a range of positions and interests, drawn from around the world. Over a period of eighteen months, OxCAIGG held expert briefings, consulted with technical advisers at the University of Oxford, produced original research into topics related to AI and good governance, and convened several days of workshops.

Based on our views as commissioners, the research to date, and our own experiences, we believe that it is time for independent and international bodies to take on the responsibility and mission of designing and supporting the implementation of standards and best practices for effective public sector application of AI.

As individual governments move forward in adopting these technologies, they would benefit from facilitating agencies that can provide guidance on engagement with industry and ways to avoid biased or otherwise inferior policy and programme outcomes, and that can offer adjudication or dispute resolution services. Such agencies would also be of benefit to many governments around the world that have limited institutional or regulatory capacity for the adoption of technology.

As a commission we prepared a series of working papers that evaluated the opportunities for and risks of the use of AI in public services and applied that learning to the development of our recommendations. Our goal is to reduce the negative outcomes of AI use and to support the development of more standardised global governance of this emerging technology.

Our goal has also been to make recommendations that are structural in nature, general in their descriptions, and based on a set of principles that are in line with our research. This approach will provide space for their flexible and innovative implementation taking into account the feedback of relevant constituencies.

As a commission we have sought to move past the development of principles for AI use—of which there are already many lists—to develop concrete policy guidelines and organisational frameworks for consideration by relevant national and international actors.

This report, *AI in the Public Service*, is not itself a literature review—such references will be found in particular research outputs produced over the life of the Commission. This is a succinct statement of our deliberations and consultations, designed to summarise what we see as the critical questions and vital objectives, our recommendations, and the next steps in the agenda for putting AI to work for good governance.

The Commissioners of the Oxford Commission on AI & Good Governance

1 Introduction

Early experiences and outcomes indicate that there will be serious challenges as governments seek to apply the tools of artificial intelligence to public service. These early days present a moment of opportunity where cooperation and consensual standards can have a significant impact in mitigating challenges and maximising opportunities.

The Oxford Commission on AI and Good Governance (OxCAIGG) has investigated several concrete challenges around AI in public service, with a specific focus on the use of AI tools for public programmes by democracies around the world. The goal of this final report of the Commission is to evaluate the challenges and to draw from core principles for the use of AI for good governance to offer specific recommendations for overcoming those challenges.

From its launch in 2020 to the production of this concluding report in 2021, OxCAIGG has:

- Funded and produced seven original research working papers, and two opinion editorials on contemporary issues surrounding AI and good governance.
- Hosted eight expert calls with stakeholders from industry, civil society, and the public sector.
- Held four round-table consultations with the OxCAIGG commissioners and technical advisers.
- Presented our work to senior policymakers in national, supranational, and multilateral public agencies.
- Submitted evidence in response to calls for expert testimony from national and multilateral public agencies.

This original research, combined with our expert briefings and engagement with public officials, has informed our deliberations and the recommendations presented here. In this report, we outline key challenges, propose a foundational set of principles, and identify the actions that should be taken now.

The proposed actions are summarised as actionable, evidence-based recommendations that will enable governments and public administrations to harness the benefits of AI. Following OxCAIGG's fundamental principles, we are committed to:

- Acting with agility and purpose to understand the impact of the AI tools currently being used in governance, in public administration, to secure social welfare, and to provide public goods and services.
- Assisting policymakers and entrepreneurs with policy questions and public applications, and in designing ideas that put AI and machine learning into public service.
- Identifying solutions and guide policymaking processes in a way that strengthens the development pathway for introducing AI to solve rather than complicate social problems and to create public trust in the use of these tools.

2 Challenges

The adoption of AI tools by the public sector combines the systemic risks inherent to AI technologies with the practical issues related to innovative public sector programming and policymaking.

In applying these tools, governments will be forced to change the way they procure, develop, and implement programmes, and they will encounter questions around the collection of training data, the assessment of complex technological systems, staff training, and the hiring of new expert staff, among other issues.

Research carried out by OxCAIGG reviewed many of the most prevalent uses of AI in governance and public service today and also looked forward, observing some challenges that will be faced in the near future. Our researchers pointed to a range of challenges that hinder the adoption of public service AI. Our recommendations seek to mitigate present challenges and provide flexibility to adapt to future ones. Here we present summaries of research undertaken by OxCAIGG and then discuss the key challenges that arose from this work.

Global Attitudes Towards AI, Machine Learning and Automated Decision Making[1]

In their working paper based on survey data from a sample of 154,195 respondents in 142 countries collected for the Lloyd's Register Foundation World Risk Poll, Lisa-Maria Neudert, Aleksi Knuutila, and Philip Howard analyse the basic indicators of public perceptions about the potential harms and opportunities of involving AI in our personal affairs and public life. They find that public perceptions of AI vary greatly across different regions and socioeconomic groups. There are notable East–West divides, with public concern about AI ranking highest in Europe (43%), Latin America (49%), and North America (47%), whereas in Southeast Asia (25%) and East Asia (11%) relatively low proportions of people thought that AI would be harmful. Across different professions, business and government executives (47%) and other professionals (44%) are most enthusiastic about AI, whereas workers in manufacturing (35%) and service workers (35%) are less confident.

Practical Lessons for Government AI Projects: Evidence from Four Smart City Initiatives[2]

In this work, Godofredo Ramizo Jr. investigates how governments employ artificial intelligence in the delivery of public services focussing on AI-driven smart city projects. Drawing from an extensive literature review and original interviews with senior government officials from Hong Kong, Malaysia, and Singapore who have worked on smart city and similar AI-driven projects, Ramizo demonstrates the diversity of government AI projects and identifies practical principles to help safeguard public interest. The working paper reveals that governments grapple with AI procurement, implementation, and impact assessment, and struggle to determine the financial, technical, and political viability of projects. In particular, when tech companies command superior resources and influence, governmental bargaining positions are challenged.

Artificial Intelligence in Local Government[3]

Thomas Vogl explores the use of artificial intelligence by UK local authorities in this working paper. While there have been a number of successful projects related to back-office automation, predictive analytics for decision support, and the use of chatbots for interactions with residents, Vogl finds that governments face important practical challenges in successfully adopting AI. He demonstrates that local authorities need to improve their capacities for data collection and analytics, more clearly define problems before seeking AI solutions, and provide suppliers with insights into contextual knowledge about local authorities and their processes.

Old Cracks, New Tech: Artificial Intelligence, Human Rights, and Good Governance in Highly Fragmented and Socially Stratified Societies. The Case of Kenya[4]

In her work, Nanjala Nyabola investigates the Kenyan government's policy on AI and blockchain technology and evaluates its success. Conducting a literature review and analysis of policy documents and working papers, she shows that in Kenya, key applications for this technology are focussed on affordable healthcare, food security, manufacturing, housing, cybersecurity, and land titling. Nyabola finds that deploying AI in highly fragmented societies like Kenya risks deepening existing cleavages, including those around class and identity, and the ethics of using AI in an industrial context versus a public-facing context differ in their implications and societal impact. Like many developing countries, Kenya is only now starting to develop legal frameworks to govern the use of technology.

Surveillance as a Service: The European AI-Assisted Mass Surveillance Marketplace[5]

Yung Au examines the European marketplace for the production and export of AI-assisted surveillance systems to governments around the world in her working paper. She investigates what she calls Surveillance as a Service—both services and software that are provided for surveillance, and which consist of complex systems that are offered with user-friendly interfaces as well as continual maintenance, updates, and troubleshooting support. Her analysis focusses on three examples of such services that have emerged as targets of major regulation as of late: facial recognition and analysis; speech recognition and analysis; and behavioural analysis and nudging systems. As the overlap between AI technologies and mass surveillance applications continues to grow, so does the potential for harm. If left under-regulated, this marketplace has widespread potential for lasting harm.

Harmonising Artificial Intelligence: The Role of Standards in the EU AI Regulation[6]

In this working paper, Mark McFadden, Kate Jones, Emily Taylor, and Georgia Osborn investigate the role of technological standards for the safe, fair, and innovative development of artificial intelligence, laid out in the draft EU AI Regulation. The working paper reveals that standardisation in this context is complicated and the nexus between standards and the European Commission's goals is a challenging intersection of stakeholders, economic interests, and established standards development organisations. Building on extensive research and stakeholder consultation, the draft regulation sets out a comprehensive framework for AI governance and standards. This paper focusses on the role that the draft regulation gives to standards for AI. Specifically, conformance with harmonised standards will create a presumption of conformity for high-risk AI applications and services—lending a level of confidence that they are in compliance with the onerous and complex requirements of the proposed regulation and creating strong incentives for industry to comply with European standards.

Key Challenges

These reports and briefing documents present original research, or present a wide array of cutting-edge research in original ways. Together, they draw from the growing amount of social and policy science about how AI is used in public service, and select from experiences and case studies around the world. They identify the key challenges for putting AI to work in the public interest. The positive and constructive examples can inspire hope, while the bad experiences deserve critique and should inform future decisions.

First, several contributions emphasise the institutional and structural complexities of setting standards for AI. While there are evident benefits to some level of standardisation, a scientifically backed roadmap for standardisation and dedicated capacities for AI standardisation are needed. Currently, there are no effective mechanisms for dispute resolution.

Second, there are very real, practical challenges surrounding the adoption of AI for public service in relation to the procurement of AI and the collection and analysis of training data. Public servants lack expertise and skills, but also practical toolkits to make good decisions. It is clear that powerful technology companies have superior bargaining power and expertise in comparison with governments and public administrators. Public servants require technical and practical capacities for the adoption of AI for good governance.

Third, a core challenge facing the use of public sector AI is public trust itself. Having public confidence in government services is always critical, and any sense that the application of machine learning is costly, undermines equity, or causes new problems will generate unnecessary roadblocks.

Finally, even well-intentioned use of AI by governments can reinforce existing biases and inequalities. A lack of legal or practical frameworks for public servants poses challenges for the adoption of these new technologies. AI technologies used for governance can have systemic biases, unforeseen consequences,

and even systematic risks to human rights. And an under-regulated marketplace can create de facto technical standards that impact public life by sustaining those biases and inequalities because no public interest intervention in those markets catches and corrects the biases and inequalities that need to be redressed.

Bringing the Research Together

Drawing from the challenges revealed in this research, and based on our expert calls and internal round-table discussions, we have identified three central deficits in the regulation of public service AI.

Who should provide guidance on public service AI? Overwhelmingly, governments recognise the potential of AI and want to use it for good governance. But the adoption of AI technologies bears highly specific practical and normative challenges that governments are not equipped to address by themselves.

How can we build public service capacities for AI for good governance? Civil servants play a vital role in the adoption of AI in public service. The informed implementation of AI requires technological skill sets and education that public agencies rarely have access to.

How can we ensure public service AI is trustworthy and trusted? Our research shows that public trust in AI is contested. We must ensure that AI systems are both trustworthy and trusted in order for them to be successfully used.

3 Principles

Today's AI landscape is full of ethical codices and normative guidelines. The world's leading experts at the world's leading institutions have proposed principles for the use of artificial intelligence in domains like criminal justice, healthcare, and sustainable development.

It is not the intention of OxCAIGG to add to the growing body of high-level mandates for AI. Rather, given the experience and expertise of this group, we have sought to develop a short and clear set of actionable recommendations that can inform the public sector use of AI. Yet from the launch of the Commission, we wanted to clearly express our mission to advance the use of AI specifically in the public service and to advance good governance. Thus, our principles balance the need for frameworks that are flexible enough to have global applicability, but precise enough to inform very specific actions taken by public servants and politicians in the application of these new tools.[7]

The work of OxCAIGG is informed by the following four key principles that formed the basis for our work as a commission and the recommendations put forward in this document:

I **Design of AI must be inclusive:**
AI tools and programmes must be informed by the experience of public sector experts to overcome challenges of discrimination and bias related to the use of inadequate data sets, the exclusion of minorities and under-represented groups, and the lack of diversity in design.

II **Procurement of AI must be transparent:**
This will overcome challenges around the acquisition and development, design, and usability of AI tools. The procurement process must include an assessment of the risks and benefits of implementing the AI tools.

III **Implementation of AI must be informed:**
Public servants need to be trained on issues of interoperability, explainability, bias, and integration with decision-making processes.

IV **AI must be accountable:**
Decisions made by an AI system must be transparent and avoid "black box" outcomes. This includes the introduction of processes to monitor and audit the AI system.

4 Recommendations

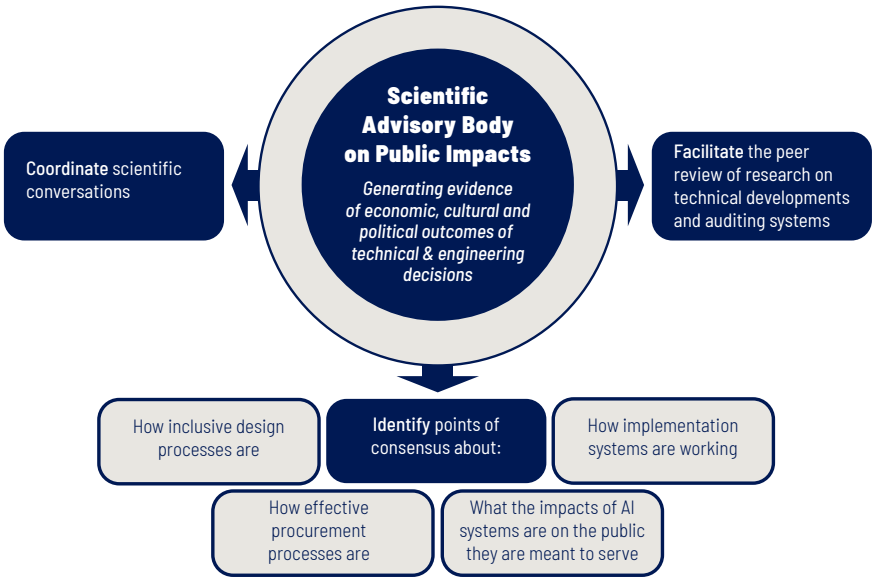
WHO SHOULD PROVIDE GUIDANCE ON PUBLIC SERVICE AI?

The Commission discussed fundamental questions around the development, procurement, and use of artificial intelligence by the public sector. Through our deliberations we identified two regulatory needs. The first is for a dedicated, international, scientific body to advance research on algorithmic audits, social impacts, use cases, and best policy practices, and to disseminate such research to inspire and coordinate the use of new AI systems for solving problems that require collective action. The second is for an arbitral body that can quickly and effectively resolve disputes between the developers, regulators, and subjects of public use AI systems. These organisations would be independent but complementary: the scientific body would provide impartial evidence to the arbitral body; the arbitral body would signal the need for research on the issues brought before it.

Scientific Advisory Body on Public Impacts

The first global body, *a scientific advisory body*, would be dedicated to scientific, engineering, and technical conversations about the impact of AI, machine learning, and other advanced algorithmic systems on public problems. Modelled on the Intergovernmental Panel on Climate Change (IPCC), such an agency would coordinate scientific conversations, facilitate the peer review of research on technical developments and auditing systems, and identify points of consensus about how inclusive design processes are, how effective procurement processes are, how implementation systems are working, and what the impacts of AI systems are on the publics they are intended to serve. Most important, however, is that this body be understood to have a technocratic but policy role. Its leadership team must be able to generate evidence of the economic, cultural, and political outcomes of technical and engineering decisions.

The scientific body should have the technical expertise to understand how artificial intelligence and social systems interact. It is an inherently multidisciplinary project that seeks to explain how AI has a concrete impact on cultural, economic, and political life around the world, and to collect and assess evidence of the impact of AI on the human rights of individuals. Operationalising inclusive design, informed procurement, purposeful implementation, and persistent accountability will require the attention of computer and social scientists.



This will be a critical body for many governments around the world that do not have their own in-country community of expert engineers and social researchers. An appropriately staffed research body would have the responsibility of evaluating the policy routes taken by member states, hosting high-level conversations about the challenges that arise as technologies evolve, evaluating the evidence of how social equality may change, and providing a forum for collecting insights and opinions from industry, civil society, and the myriad other public interest groups impacted by AI in public service.

Critically, this scientific body will bring evidence to bear on the formation of standards and certification systems and toolkits for policymakers. The work of auditing algorithms, interpreting impacts, and evaluating best practices will need to be led by computer, social, and policy researchers to be credible and impactful.

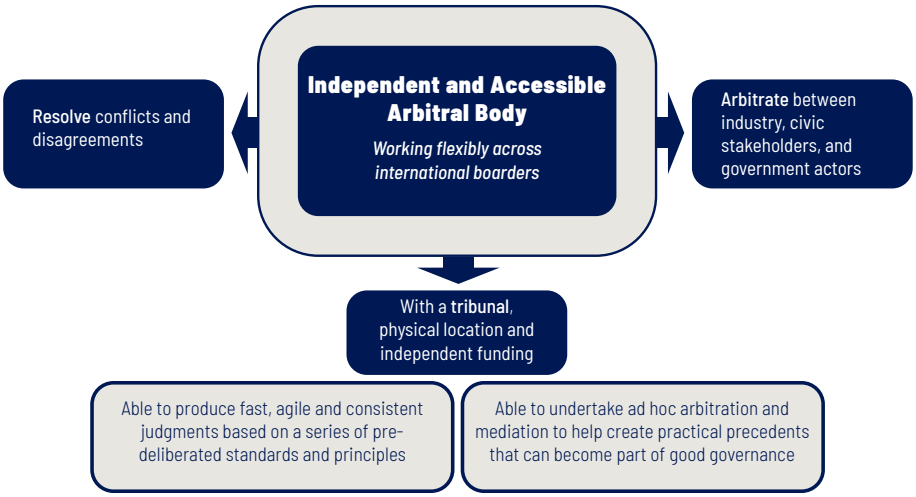
There are several kinds of international academic associations, multilateral research bodies, and intergovernmental agencies that already have research units working on some of these questions. However, coordinated scientific consensus-building around evidence of the global impact of public service AI is not systematically organised. Fortunately, the IPCC does provide a role model for how scientific learning can be coordinated. There is certainly research on how major domains of social inquiry calcify over time, but we know enough about encouraging diversity and creativity in social inquiry to mitigate against this. Moreover, the output from such a scientific body would be critical to the recommendations ahead: solid evidence and scientific consensus will be required for use by the other bodies and processes we recommend here.

Independent and Accessible Arbitral Body

The second body, *a global arbitration body*, would help resolve conflicts and disagreements, arbitrating between industry, civic stakeholders, and government actors as critical issues arise. As a commission we have observed that the pace of innovation is incredibly rapid. Trying to anticipate how new AI tools will be used in public life is difficult. By virtue of their market share, industry can set de facto technical standards, while legislation lags behind. We can be reasonably confident that disagreements and misapplications will occur, and that both technical errors and design flaws will complicate how machine learning is used in public life. Formally, some countries have legal routes through which civil action can be pursued. But many do not, many such disagreements arise across international borders, and the vast majority of national arbitral courts have limited technical capacity to evaluate disputes. What is more, the pace of innovation has proven to outpace even supposedly future-proof legislation.

A dedicated arbitral body with a high-capacity secretariat—with a physical location, independent funding, and the ability to provide arbitration and mediation services—would help create practical precedents that can become part of good governance. The mandate of the arbitral tribunal would be to produce fast, agile, and consistent judgements based on a series of standards and principles that would, presumably, be suggested by the scientific body. The speed of technological development today makes reliance on wide-ranging, and sometimes conflicting, national laws practically impossible. Globally we need flexible arbitration around real-time issues that arise as the technology evolves.

Clearly the specifics of how the arbitral body will work has to be negotiated, perhaps as one of the starting concerns for the



scientific body. The panel of judges will need to be diverse. It should have a formal seat or location in a stable, politically neutral jurisdiction with a strong legislative respect for arbitral decisions. It will need a secretariat to assist with collating evidence and submissions from interested parties. It will need a funding mechanism—ideally a mixture of support from government and industry with appropriate firewalls in place. It will be critical that the arbitral body be able to receive notes from a wide constituency of stakeholders—similar to the role that amicus briefs play in some national court systems. This will allow stakeholders that are not a party to a disagreement to assist the arbitration process by offering information, expertise, and insight that has a bearing on the issues under consideration.

The tribunal will adjudicate disputes or disagreements between and among the public, private, and civil sectors. These may arise, for example, around intellectual property rights, the scope of public mandates for the supply of services, or the consequences of AI use by public agencies. In our discussions, the Commission resolved that the arbitration panel would need to have authority through a contractual or declaratory "opt in" from both public and private sector actors, who would see the advantages of transparent standards, credible evaluations, rapid response, and globally consistent rulings. It will be critical that both sectors agree to be bound by the decisions of the arbitral tribunal, meaning that governments would have to waive claims of sovereign immunity and agree to be bound by the rules and decisions within the boundaries of national laws. Similarly, private actors would have to agree to abide by the decisions of the arbitral body, and not pursue additional relief or alternative outcomes through appeals to other less competent or relevant bodies.

The advantage of an arbitral court is that it generates actively useful precedent decisions, often flexibly in response to contemporary issues, and can move more quickly than the processes that generate detailed law. And a dedicated arbitral court for the use of AI in public administration would provide a special depth of commitment and technical capacity for resolving disagreements.

Working Together

AI has become prolific, but its use for good governance lacks mechanisms for inclusion, collaboration, and participation among government, industry, civic stakeholders, and researchers. More transnational collaboration across the scientific and arbitral bodies could foster trust in AI regulation, increase consensus, and ultimately bolster civic support for how AI is implemented and used in governance. Demanding "more collaboration" is hardly sufficient. Collaboration should be meaningful, engaging international stakeholders at arm's length and in a dialogue- and consensus-oriented but inclusive and targeted manner.

Joint policy initiatives and goals on the European or transatlantic level, industry collaboration, scientific conferences, and funding for AI-focussed research and advocacy projects are offering a starting point for facilitating this kind of networking and knowledge exchange. But geopolitical differences and competitive pressures have hindered arm's-length collaboration across different stakeholder groups and national borders. Independent arbitral systems and scientific consensus-building bodies can help advance collaboration—or at the very least, dialogue—between government, industry, and civil society, nationally and internationally.

Altogether, this set of organisations, if well resourced, would provide comprehensive guidance as innovations occur in the years ahead. They would function well as a set, with one dedicated to generating evidence using the latest scientific methods and the other dedicated to evaluating social consequences using such evidence. There is good precedent in the international system for organising this way. Arbitral decisions are enforceable through the 1958 New York Arbitration Convention (Convention on the Recognition and Enforcement of Foreign Arbitral Awards), to which 168 countries are parties. The Court of Arbitration for Sport has a strong international reputation for speedy decision-making and technical prowess, and has also earned widespread credibility for its work. The International Council for Advertising Self-Regulation and International Chamber of Commerce are both well-established organisations that set standards, offer private, binding arbitration, and have consultative status with the United Nations. Certainly, more research would be needed to establish the legal framework through which industry, government, and civil society would collaborate—likely through private law. But the role models for how such collaboration can occur are encouraging.

RECOMMENDATION 1

Internationally, governments, industry, and civil society must work together to build and empower (a) an international scientific body for advancing research on public service AI applications and (b) an arbitral body for adjudicating disputes that might arise between the stakeholders involved with public service AI systems.

HOW CAN WE BUILD PUBLIC SERVICE CAPACITIES FOR AI FOR GOOD GOVERNANCE?

AI, machine learning, and other advanced algorithmic systems have evolved extremely rapidly, and we should anticipate that they will continue to do so. AI encompasses virtually all areas of public life and thus has emerged as a cross-sectional challenge for regulation. We have proposed bodies to devise principles and standards and have sight of compliance. But questions around the everyday use of AI in public service are often far more mundane and of a practical nature. What is needed to equip public servants to effectively regulate AI in public service? Here, we provide a broad conception of regulation, including administrative rule-making within the mandate of principles or standards, such as those devised by the proposed scientific body.

Firstly, we recommend applying the principles we have enunciated in section 3, and we recommend regulation around the *design* of AI systems; the *procurement* processes by which public agencies acquire or license machine learning systems; the *implementation* processes, including public consultations, awareness campaigns, and information access; and the long-term *accountability* processes for collating public feedback about the long-term consequences of applied machine learning on collective action problems.

Many countries have already successfully adopted AI in the public service—with or without regulation around design, procurement, implementation, and accountability. But our research stresses that public servants still widely lack the capacities needed for the procurement, design, evaluation, and implementation of AI. The next step is to provide public servants with fundamental expertise to capacitate them to make good decisions on AI. Certainly, education of public servants and upskilling of regulatory agencies are central to this challenge, but these processes are resource- and time-intensive. In the interim, practical toolkits providing best practices and simple decision matrices may remedy some of the most pressing challenges. It is essential that governments create a central and accessible knowledge base, so that proven strategies and expertise can flow across different departments.

In order to do so, it is vital to equip regulators with the competencies and mandate to develop the consultative instruments and compliance mechanisms they require. Certainly, some parts of the regulatory work for AI design, procurement, implementation, and accountability might be outsourced—not all of it need be done by governments. It is likely that public agencies should lead on the first three areas, with accountability and certification work done by third parties under some public guidelines.

Finally, while it is useful to identify *what* should be regulated, it is important to note here that the process of regulation—and any accompanying regulatory programme—must have sufficient resources and capacities to support teams of intergovernmental staff who can specialise in public sector AI issues. It will be vital to keep the administrators of public programmes well trained, but it should also be possible to extend technology literacy across government agencies in ways that equip governments more broadly with the ability to interact with other industry and civil society stakeholders.

RECOMMENDATION 2

National governments, taking advantage of support from these two new international organisations, must (a) build the capacity of their public service to be deeply engaged with the design, procurement, implementation, and accountability of public service AI systems and (b) provide toolkits for the staff of public agencies to do their supervisory work.

HOW CAN WE ENSURE PUBLIC SERVICE AI IS TRUSTWORTHY AND TRUSTED?

A lack of trust could become a critical barrier to the successful and timely implementation of AI in the public service: if the public does not trust AI, political decision-making based on AI systems is bound to face vehement opposition. And certainly, public trust in AI is only desirable when technical systems are indeed trustworthy. Therefore, OxCAIGG argues that both bolstering public support and AI assessments are essential to the implementation of AI for good governance. We recommend that governments and public agencies address civic worries about the use of AI and address them with measures aimed at education, literacy, and certification.

Firstly, to strengthen public trust in public service AI, we recommend that governments launch public communication and literacy campaigns that highlight benevolent uses of AI for public interest issues in domains with which the public is already familiar. Communication campaigns should also take into account valid points of concern surrounding systemic flaws and government misuse of AI systems. We recommend that governments commit substantial communication budgets to timely, cross-media campaigns and marketing, aiming to reach various demographic groups. As negative narratives and misinformation about AI have become prevalent, shedding a light on the use of AI in areas like early diagnostics, traffic management, and climate tech could support civic approval of AI technologies.

Secondly, to inform further work in this area in the longer term, governments should strive to better disclose when and how they use AI or plan to use it. Disclosing where and how AI is used in public service would serve to increase transparency and ultimately trust in AI. Disclosure notes on government web pages and central databases, as well as publicly accessible lists of AI suppliers and government-developed AI systems, could serve as a starting point towards comprehensive disclosure systems. There is benefit in a consolidated, international database that could be put forward by the scientific body we have recommended.

Thirdly, we propose a global certification system for the use of AI in public service. Such AI certification should take into account both safety and quality indicators, and keep a public record of government use of AI systems that has been deemed insufficient. The scholarly debate on the operationalisation of benevolent and trustworthy AI is well evolved and researchers have put forward normative ethical frameworks and measurable indicators to assess the transparency, explainability, and accountability of AI systems. As a result, quality, risk, and impact assessments, rigorous testing in design processes, training data evaluation, and system maintenance have become more widely adopted in the industry, but comprehensive best practices or standards are yet to form. There are good reasons not to fear that certification will discourage innovation. Around the world there are capable audit and certification practices for financial trading software, video gambling machines, and data processing. Devising comprehensive certification systems and policy toolkits may be one of the first mandates of the proposed scientific body, though an international standards agency or industry-based agency may eventually be best placed to pursue certification.

RECOMMENDATION 3

Trust in the use of AI must be strengthened through public education campaigns about the everyday applications and real use cases on the horizon, their impacts, and their risks, by (a) having governments openly disclose how AI technologies are being used in the public service and (b) introducing a multisector agency which provides a basic certification system that continually validates applications and builds trust over time.

5 Conclusion: Ideas for Immediate Action

Identifying broad principles is critical for providing a framework for understanding a public problem. It has helped us evaluate challenges and identify concrete recommendations for action.

Having developed the three recommendations in this report now allows us to signal what the next steps should be. Knowing what we know about the challenges of putting AI into public service, and equipped with a set of recommendations for how to support good governance through AI systems, what should be the next steps?

I. AN ORGANISATIONAL FEASIBILITY STUDY

A feasibility study would allow us to estimate the costs of organising the arbitral and scientific bodies that are needed. Clearly these agencies will need to be "seated" in legal jurisdictions and political climates conducive to their operations. Who would lead such organisations, and what specific organisational capacities will be needed to see through their vital missions? The first step to take is to prepare a feasibility study that can fill in the logistical and operational scenarios for building our global capacity to put AI into public service.

II. CONSULTATION WITH EXISTING NATIONAL AND MULTILATERAL AGENCIES

There are a number of high-capacity, multilateral institutions that are making important contributions to the global conversation about AI in public service. No single one of these agencies has an extensive purview, however, or treats this work as their central, dedicated mission. Moreover, several aspects of the work that needs to be done fall between the mission scopes of existing multilateral agencies. Experts at UNESCO have sight of many of the important trends in AI applications, and the United Nations Development Programme and both the G7 and the G20 have begun discussing the role of AI in governance. The UN's Inter-Agency Working Group on AI meets regularly, and is being led by UNESCO and the International Telecommunication Union.

We argue that global policymakers and decision makers are the ones who need to pave the way towards more meaningful collaboration. The G7's Global Partnership on Artificial Intelligence (GPAI) is a laudable effort in this direction, bringing together experts from science, technology, civil society, and policy in a multi-stakeholder manner. But the GPAI is a fairly exclusive group and lacks representation from some of the most important innovators in AI, as well as from the so-called Global South. Notably China, a global leader on AI, has remained disconnected from the wider discourse and concerns about human rights infringements relating to the use of AI. Based on our research and expert discussions at OxCAIGG, we stress the importance of more international collaboration and exchange.

III. INCLUSIVE CONVERSATIONS AT INTERNATIONAL MILESTONE EVENTS

There are a number of important occasions on the horizon—events at which inclusive conversations can be held about the realistic prospects for improving our capacity to ensure that AI is used for good governance. For example, in September 2023, the UN will hold the Summit of the Future, bringing together global policymakers at the head of state level. This event may serve as a locus for dialogue on neutral territory and a starting point for more continuous exchange—including among stakeholder groups that share marginal consensus, if any. A global event of that calibre may be well placed as a launch pad for more structured and long-term collaboration.

Objectives Summary

RECOMMENDATIONS

1

Internationally, governments, industry, and civil society must work together to build and empower
(a) an international scientific body for advancing research on public service AI applications and
(b) an arbitral body for adjudicating disputes that might arise between the stakeholders involved with public service AI systems.

2

National governments, taking advantage of support from these two new international organisations, must
(a) build the capacity of their public service to be deeply engaged with the design, procurement, implementation, and accountability of public service AI systems and
(b) provide toolkits for the staff of public agencies to do their supervisory work.

3

Trust in the use of AI must be strengthened through public education campaigns about the everyday applications and real use cases on the horizon, their impacts, and their risks, by
(a) having governments openly disclose how AI technologies are being used in the public service and
(b) introducing a multisector agency which provides a basic certification system that continually validates applications and builds trust over time.

NEXT STEPS FOR IMMEDIATE ACTION

I

An organizational feasibility study

II

Consult with existing national and multilateral agencies

III

Inclusive conversations at international milestone events

6. OxCAIGG Research

[1] Neudert, L.-M., Knuutila, A. & Howard, P. N. *Global Attitudes towards AI, Machine Learning & Automated Decision Making*. (Working Paper 2020.10, Oxford Commission on AI & Good Governance, 2020).

[2] Ramizo, G., Jr. *Practical Lessons for Government AI Projects*. (Working Paper 2021.1, Oxford Commission on AI & Good Governance, 2021).

[3] Vogl, T. *Artificial Intelligence in Local Government*. (Working Paper 2021.2, Oxford Commission on AI & Good Governance, 2021).

[4] Nyabola, N. *Old Cracks, New Tech: Artificial Intelligence, Human Rights and Good Governance in Highly Fragmented and Socially Stratified Societies*. *The Case of Kenya.* (Working Paper 2021.3, Oxford Commission on AI & Good Governance, 2021).

[5] Au, Y. *Surveillance as a Service: The European AI-Assisted Mass Surveillance Marketplace*. (Working Paper 2021.4, Oxford Commission on AI & Good Governance, 2021).

[6] McFadden, M., Jones, K., Taylor, E. & Osborn, G. *Harmonising Artificial Intelligence: The Role of Standards in the EU AI Regulation*. (Working Paper 2021.5, Oxford Commission on AI & Good Governance, 2021).

[7] Neudert, L.-M. & Howard, P. N. *Four Principles for Integrating AI and Good Governance*. (Working Paper 2020.1, Oxford Commission on AI & Good Governance, 2020).

About the Commission

The challenge of using AI for good governance and public service is an urgent concern in countries across the world. Launched in July 2020, the goal of the Oxford Commission on AI and Good Governance has been to develop principles and practical policy recommendations to ensure the democratic use of AI for good governance.

Most recently, the COVID-19 pandemic has prompted a rapid influx of AI solutions. While intended for public good, these novel technologies bring with them challenges in assessing the suitability and legitimacy of these offerings. The rapidity of implementation of such systems is unprecedented and demonstrates the need for policies around these sorts of AI products, their procurement, and their implementation through governments.

OxCAIGG investigated the procurement and implementation challenges surrounding the use of AI for good governance that are faced by democracies around the world, identified best practices for evaluating and managing risks and benefits, and recommended strategies in an effort to take full advantage of technological capacities while mitigating potential harms of AI-enabled public policy.

Drawing from input from experts across a wide range of geographic regions and areas of expertise, including stakeholders from government, industry, and technical and civil society, OxCAIGG produced applicable and relevant recommendations for the use of AI for good governance.

Our commissioners applied their experience and insight to contribute thoughtfully to OxCAIGG's quest to equip and inform policymakers with guidance to ensure AI-related tools are adapted and adopted for good governance in the near future.

Commissioners

Dr Yuichiro Anzai, OxCAIGG Commissioner

Senior Advisor for the Japan Society for the Promotion of Science and Executive Advisor for Academic Affairs at Keio University.

Yuichiro Anzai is engaged in giving advice, as the chair of the Council for Artificial Intelligence Strategy, to the Japanese government for making strategic policies. He was awarded the title Person of Cultural Merit by the Japanese government for his pioneering work on the integration of cognitive and information sciences.

Dr Rumman Chowdhury, OxCAIGG Commissioner

Director of the META (ML Ethics, Transparency, and Accountability) team at Twitter.

Rumman Chowdhury's passion lies at the intersection of artificial intelligence and humanity. She is a pioneer in the field of applied algorithmic ethics, creating cutting-edge socio-technical solutions for ethical, explainable, and transparent AI.

Mr Tom Fletcher (CMG), OxCAIGG Commissioner

Principal of Hertford College, University of Oxford, and founder of the Foundation for Opportunity.

Tom Fletcher was previously the foreign policy adviser to three UK prime ministers; the UK's Ambassador to Lebanon; a visiting professor at New York University; an adviser to the Global Business Coalition for Education; and chaired the International Board of the Creative Industries Federation. He is the author of The Naked Diplomat.

Professor Dame Wendy Hall (DBE, FRS, FREng), OxCAIGG Commissioner

Regius Professor of Computer Science, Associate Vice President (International Engagement), and Executive Director of the Web Science Institute at the University of Southampton.

Wendy Hall is Chair of the Ada Lovelace Institute and a member of the BT Technology Advisory Board. She became a Dame Commander of the British Empire in the 2009 UK New Year's Honours list, and is a Fellow of the Royal Society.

Professor Philip Howard, OxCAIGG Commissioner

Statutory Professor of Internet Studies, Balliol College, University of Oxford.

Philip Howard investigates the impact of digital media on political life around the world, and he is a frequent commentator on global media and political affairs. He is Director of Oxford University's Programme on Democracy and Technology, which investigates the use of algorithms and automation in civic life.

Sir Julian King, OxCAIGG Commissioner

Former European Commissioner and British diplomat.

Julian King previously served as UK Ambassador to Ireland and France and Director General of the Northern Ireland Office, and was the last British official to hold a position and portfolio within the European Commission, as Commissioner for the Security Union.

Dr Safiya Noble, OxCAIGG Commissioner

Associate Professor at the University of California, Los Angeles (UCLA) in the Department African American Studies and the Department of Information Studies and Co-Director of the UCLA Center for Critical Internet Inquiry (C2i2).

Safiya Noble's work is both sociological and interdisciplinary, marking the ways that digital media impacts and intersects with issues of race, gender, culture, and technology. She is a board member of the Cyber Civil Rights Initiative, serving those vulnerable to online harassment.

Mr Howard Rosen (CBE), OxCAIGG Commissioner

Principal of Howard Rosen Solicitors, President of the Swiss Friends of Oxford University, Chairman of the Rail Working Group, and Trustee of the Commonwealth Jewish Council.

Howard Rosen specialises in international commercial, finance, and leasing law and trusts. He is also the founder and Managing Director of Rosetrust AG, a trust company in Zug, Switzerland, and of Aviation Advocacy Sàrl, based in Nyon, Switzerland.

Professor Weixing Shen, OxCAIGG Commissioner

Dean of Tsinghua University's School of Law in Beijing, Director of the Institute for Studies on AI & Law, and Faculty Director of the LLM Program on Computational Law of Tsinghua University.

Weixing Shen is Commissioner of the Artificial Intelligence Industry Alliance in China and also serves as an Executive Council Member of China's Law Society and as Vice President of the Cyber and Information Law Association of China's Law Society.

Baroness Joanna Shields (OBE), OxCAIGG Commissioner

Chief Executive Officer, BenevolentAI

Joanna Shields is a tech industry veteran with a successful track record building some of the world's best-known companies. She is extremely passionate about tech that benefits humanity. Her career spans over 30 years and has focussed on harnessing the power of technology to drive change that improves connectivity, humanity, and society.

Acknowledgements

We are grateful to Flora Seddon for coordinating and supporting our OxCAIGG activities over the past 18 months. For support with our digital consultations and strategic conversation we are grateful to Bruno Selun and the team at Kumquat. For their contributions, we are thankful to Hubert Au, Rutendo Chabikwa, Dr. Tim Curnow, Mona Elswah, John Gilbert, Mark Healy, Dr. Lucy Hennings, Mark Malbas, Dr. Nahema Marchal, Sara Spinks, and Niamh Walsh. We thank the experts, researchers, and public service specialists who have contributed to this report and OxCAIGG.

For supporting OxCAIGG we are grateful to the Adessium Foundation, Civitates, the Ford Foundation, Luminate, and the Open Society Foundations. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the Commission and do not necessarily reflect the views of the University of Oxford, our funders, or individual commissioners. Ethical oversight of research at the University of Oxford is done by its Central University Research Ethics Committee, and approval numbers for individual research outputs can be found within each working paper or report, as appropriate.

oxcaigg.oii.ox.ac.uk

Oxford Commission on AI & Good Governance.
AI in the Public Service: From Principles to Practice.
Working paper 2021.6 Oxford, UK:
Oxford Commission on AI & Good Governance. 19 pp.

Retrieved from: https://oxcaigg.oii.ox.ac.uk